The EMBO Journal (2004) 23, 3196–3205 | © 2004 European Molecular Biology Organization | All Rights Reserved 0261-4189/04
www.embojournal.org

THE EMBO JOURNAL

Crystal structure of glycogen synthase: homologous enzymes catalyze glycogen synthesis and degradation

Alejandro Buschiazzo[1], Juan E Ugalde[2], Marcelo E Guerin[1], William Shepard[3], Rodolfo A Ugalde[2] and Pedro M Alzari[1,*]

[1]Unité de Biochimie Structurale, URA 2185 CNRS, Institut Pasteur, Paris, France, [2]Instituto de Investigaciones Biotecnológicas, Universidad de General San Martín and CONICET, CC 30, San Martín, Argentina and [3]European Synchrotron Radiation Facility, Grenoble, France

Glycogen and starch are the major readily accessible energy storage compounds in nearly all living organisms. Glycogen is a very large branched glucose homopolymer containing about 90% α-1,4-glucosidic linkages and 10% α-1,6 linkages. Its synthesis and degradation constitute central pathways in the metabolism of living cells regulating a global carbon/energy buffer compartment. Glycogen biosynthesis involves the action of several enzymes among which glycogen synthase catalyzes the synthesis of the α-1,4-glucose backbone. We now report the first crystal structure of glycogen synthase in the presence and absence of adenosine diphosphate. The overall fold and the active site architecture of the protein are remarkably similar to those of glycogen phosphorylase, indicating a common catalytic mechanism and comparable substrate-binding properties. In contrast to glycogen phosphorylase, glycogen synthase has a much wider catalytic cleft, which is predicted to undergo an important interdomain 'closure' movement during the catalytic cycle. The structures also provide useful hints to shed light on the allosteric regulation mechanisms of yeast/mammalian glycogen synthases.
The EMBO Journal (2004) **23,** 3196–3205. doi:10.1038/sj.emboj.7600324; Published online 22 July 2004
Subject Categories: structural biology; cellular metabolism
Keywords: glycogen; glycosyltransferase; starch 3D structure; X-ray crystallography

Introduction

Glycogen and starch, megadalton-sized glucose polymers, are the major reservoir of readily available energy and carbon compounds in most living organisms, ranging from archaea, eubacteria and yeasts, up to higher eukaryotes including plants and animals. Only parasitic lifestyles seem to be related to reduction and eventual complete abolishment of glycogen metabolism (Henrissat *et al*, 2002). In mammals, glucose uptake and utilization are under tight control. Defects in normal glucose handling are associated with a variety of human pathologies, like glycogen storage diseases (Roach, 2002), and diabetes, in which persistent hyperglycemia is correlated with early onset and severity of the disease (Saltiel, 2001). In the field of diabetes, several lines of evidence support a direct link between glycogen metabolism and overall glucose flux into the cells. Transgenic mice overexpressing an activated form of glycogen synthase (GS) showed glycogen overaccumulation (Manchester *et al*, 1996; Azpiazu *et al*, 2000), a phenotype that was also observed in mice overexpressing the phosphatase targeting subunit R_{GL}, which caused hyperactive GS through extensive dephosphorylation (Aschenbach *et al*, 2001). The fact that GS activity correlates with glycogen levels strongly suggests that membrane glucose transporters (GLUT) are not saturated, highlighting the importance of glycogen metabolism regulation in glucose handling.

The basic metabolic pathways for the catalytic conversion of glycogen involve the action of several enzymes, among which GS catalyzes the elongation, and glycogen phosphorylase (GP) the breakdown, of α-(1,4)-linked glucans (Figure 1). GS is a key component of this enzymatic machinery, catalyzing the successive addition of α-1,4-linked glucose residues to the nonreducing end of glycogen, using nucleoside-diphospho-glucose as the donor substrate (Figure 1A). The other components of this machinery are GP, which functions as a depolymerizing enzyme (Figure 1B), and branching/debranching enzymes that catalyze the addition and removal of α-1,6-linked ramifications. While GS and GP are glycosyltransferases (GTs), branching and debranching enzymes are variants of glycosylhydrolases (GHs) with acquired transglycosylation activities. GTs and GHs have been classified according to their sequence, generating a database with high functional predictive power (see http://afmb.cnrs-mrs.fr/CAZY/; Coutinho *et al*, 2003). Following the conventions of the CAZY classification, GSs belong to families GT3 and GT5, while GPs can all be grouped in GT35. GT3, GT5 and GT35 are all GTs that retain the anomeric stereochemistry among substrates and products, a property shared by at least 18 of the 70 GT families. Both branching and debranching enzymes pertain to GH13, with sequence and structural homology to α-amylase $(\beta/\alpha)_8$-fold proteins (Abad *et al*, 2002; Hondoh *et al*, 2003).

GP catalyzes a reversible reaction, which at physiological concentrations of reactants constitutes the first step of glycogen degradation, removing one glucose at a time by the addition of inorganic phosphate rendering glucose-1-phosphate (Glc-1-P) (Figure 1B). A great wealth of information has been gained on glycogen degradation through the biochemical and structural studies of GP and maltodextrin phosphorylase (MalP) in the last 30 years (Johnson, 1992; Rath *et al*, 2000). GP regulation is still a fascinating example of covalent and noncovalent allosterism, exquisitely

*Corresponding author. Unité de Biochimie Structurale, Institut Pasteur, 25 rue du Docteur Roux, 75724 Paris cedex 15, France.
Tel.: +33 1 45 68 8607; Fax: +33 1 45 68 8604;
E-mail: alzari@pasteur.fr

Received: 17 March 2004; accepted: 21 June 2004; published online: 22 July 2004

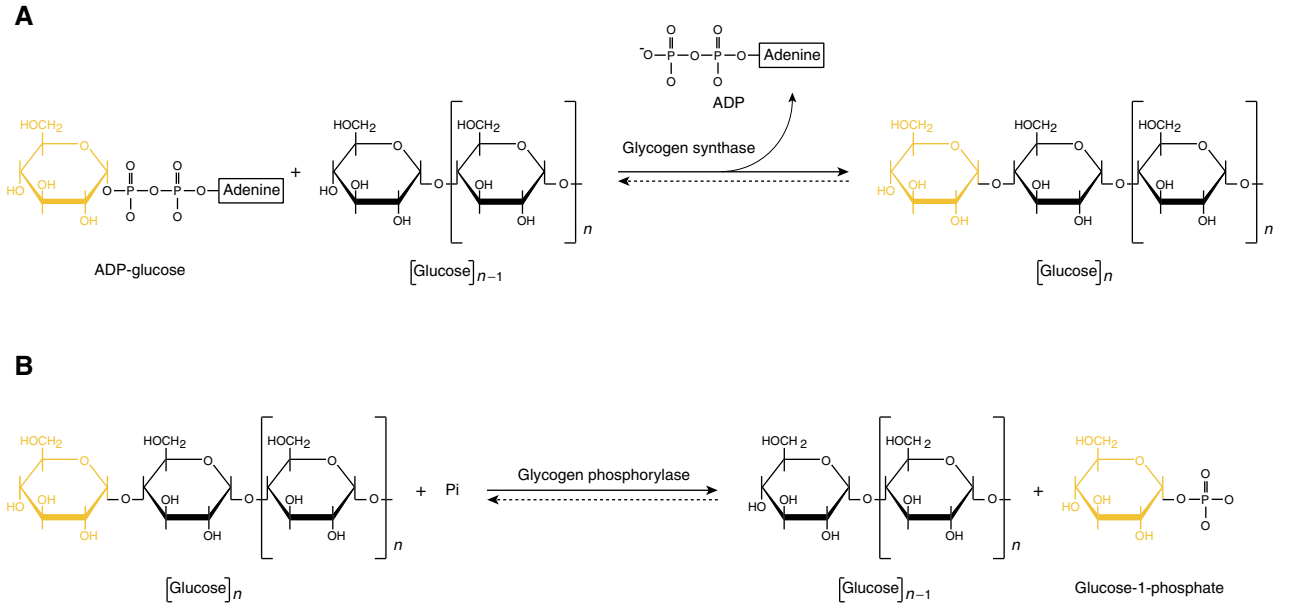


Figure 1 Glycogen synthesis and degradation. (**A**) The synthesis of glycogen involves primarily the action of GS, which catalyzes the addition of α-1,4-linked glucose to the nonreducing end of the growing glycogen chain, and the branching enzyme (not shown), which introduces α-1,6-linked ramifications. In fungi and animal cells, autoglucosylation of glycogenin (Gibbons *et al*, 2002) provides the initial glycogen primer for further elongation. (**B**) Glycogen breakdown involves the action of GP, which catalyzes the phosphorolysis of α-1,4 glucose linkages, and debranching enzyme (not shown), which hydrolyzes α-1,6 linkages. In fungi and animal cells, both GS and GP are regulated by allosteric ligand binding and phosphorylation.

coordinated by hormonal control (Buchbinder *et al*, 2001). The catalytic core of GP was rapidly found to have the same topology as the phage T4 beta-glucosyltransferase (β-GT) (Holm and Sander, 1995) defining a GT-B-fold superfamily. In fact, only two different topologies have been found among the first 12 GT families for which 3D structures have been reported (Coutinho *et al*, 2003). These topologies have been identified as GT-A (Charnock and Davies, 1999) and GT-B (Vrielink *et al*, 1994). Although both folds are variations of Rossmann-like β–α–β domains, in the GT-A fold a unique central sheet of 7–8 β-strands is found, with the presence of a DxD motif participating in ribose/metal binding in the catalytic center. In contrast, GT-B proteins do not bind metals and have two distinct Rossmann domains separated by a large cleft that permits important flexibility. Sequence similarities are marginal among GT-A or GT-B proteins, which is consistent with the fact that each fold superfamily includes a wide variety of inverting and retaining GTs, and even enzymes with no transferase activity at all.

Glycogen and starch synthases are classified in two large protein families: mammalian and yeast GSs (GT3 family) are ∼80 kDa enzymes that use UDP-glucose as sugar donor and are regulated by covalent phosphorylation and allosteric ligand binding (Roach, 2002), whereas bacterial and plant synthases (GT5 family) have a smaller size (∼50 kDa), prefer ADP-glucose and appear to be unregulated proteins (Ball and Morell, 2003). Sequence similarity is marginal between the ADP-glucose- and UDP-glucose-requiring enzymes.

Information describing the tertiary and quaternary structures is available for all the enzymes directly involved in glycogen processing, except for GS. Although several groups had predicted a general relatedness among GTs displaying the GT-B fold, including GSs (Wrabl and Grishin, 2001; MacGregor, 2002; Yep *et al*, 2004), their suggestions were based on sequence analyses with marginal identity levels. To further advance in the understanding of glycogen synthesis, we determined the crystallographic structure of the GS from *Agrobacterium tumefaciens* (family GT5), in the presence and absence of ADP. The initial suggestion that glycogen synthesis simply represented a reversal of its degradative phosphorolysis was demonstrated to be flawed, as Leloir and Cardini (1957) showed that UDP-Glc and GS, not Glc-1-P and GP, were responsible for polysaccharide elongation. The present communication demonstrates nevertheless that the opposite independent reactions in glucose–glycogen interconversion are catalyzed by closely related homologous enzymes.

Results and discussion

Overall structure

The crystal structure of *A. tumefaciens* GS (AtGS), both alone and in the presence of ADP, was determined at 2.3 Å resolution (Table I). The enzyme folds into two Rossmann-fold domains (Figure 2A) organized as in GP and other GTs of the GT-B superfamily (Coutinho *et al*, 2003), with a deep fissure between both domains that includes the catalytic center. The core of the N-terminal domain (residues 1–244) consists of a nine-stranded, predominantly parallel, central β-sheet (Figure 2B) flanked on both sides by seven α-helices. The C-terminal domain (residues 271–456) shows a similar fold with a six-stranded parallel β-sheet and nine α-helices. The last α-helix of this domain undergoes a kink at position 457–460, with the final 17 residues of the protein (461–477) crossing over to the N-terminal domain and continuing as α-helix, a typical feature of GT-B enzymes (Wrabl and Grishin, 2001).

Table I Data collection, phasing and refinement statistics

Data set	SeMet	ADP complex
Wavelength (Å)	0.9792	0.9393
Resolution (Å)[a]	30–2.3 (2.4–2.3)	50–2.3 (2.42–2.3)
Measured reflections	303 988	73 895
Multiplicity[a]	6.7 (3.8)	1.8 (1.8)
Completeness (%)[a]	98.2 (85.9)	90.7 (85.5)
R_{sym} (%)[a, b]	5.5 (18.8)	6.2 (24.9)
I/σ[a]	11.4 (3.9)	6.9 (3)
a, *b*, *c* (Å)	69.3, 87.3, 88.5	69.35, 88.1, 84.45
β (deg)	100.3	98.2
Mean FOM acentric	0.33	
Anomalous phasing power[c]	1.039 overall	
	1.5 3.7–3.2 Å	
	1.066 3.2–2.9 Å	
R_{cryst}[d] (%) [no. of reflect.]	20.4 [84 210]	18 [36 316]
R_{free}[d] (%) [no. of reflect.]	24 [4397]	22.3 [4055]
R.m.s. bonds (Å)	0.009	0.019
R.m.s. angles (deg)	1.7	1.8
Protein atoms	7246	7234
Water atoms	409	207
ADP atoms	—	54

Data were collected on beamline ID29 at the European Synchrotron Radiation Facility (ESRF, Grenoble, France).
[a]Values in parentheses apply to the high-resolution shell.
[b]$R_{sym} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$.
[c]Anomalous phasing power = $\langle [|F(h)_{calc}|$/phase-integrated lack of closure$] \rangle$.
[d]$R = \sum_{hkl}|F(h)_{obs} - F(h)_{calc}| / \sum_{hkl}|F(h)_{obs}|$. R_{cryst} and R_{free} were calculated from the working and test reflection sets, respectively.

The deep and wide cleft in AtGS suggests that the enzyme crystallized in a relaxed, possibly inactive, 'open' state. The hypothesis of a closure movement can be raised, supported by the structural superpositions of N- and C-terminal domains from several GT-B enzymes (see below). These comparisons show that a subdomain rotation of 20–25° is required to achieve a 'closed' state (Figure 2C), which would bring together critical conserved residues making up a functionally competent active center. Although ADP does bind to the open GS state in the crystal inducing minor subdomain rearrangements, soaking of either ADP-Glc, or ADP and maltodextrins, results in crystal dissolution, suggesting that substrate binding may promote the conformational transition. Moderate flexibility of such two-domain enzymes is not rare (Vrielink *et al*, 1994; Morera *et al*, 2001; Hu *et al*, 2003). Indeed, the four crystallographically independent molecules of AtGS in its free and ADP-bound forms show variations in the interdomain orientation of up to 6° (Figure 2D), comparable to those observed in the allosteric transition of GP (Buchbinder *et al*, 2001).

ADP-binding pocket

In the open form of AtGS, ADP binds to a pocket on the C-terminal side of the interdomain crevice (Figure 3). This ADP-binding pocket (Figure 3A) is defined by the C-terminal ends of β-strands β13 (residues 297–299) and β14 (326–328) and the loop between β15 and α12 (352–356). The adenine heterocycle is stabilized by a stacking interaction with Tyr354 and additional van der Waals interactions with Ser298 and Ser359 (Figure 3B). Only two weak hydrogen bonds are observed for this part of the ligand, between atoms N1 and N6 of the adenine heterocycle with the amide and carbonyl groups of the protein backbone at positions 353 and 355. The ribose O2 atom makes hydrogen bonds with the main-chain carbonyl group of Ile297 and the side chain OH of Thr381, but otherwise the sugar ring is relatively free of contacts in the crystal structure. The phosphate groups make hydrogen-bonding interactions with a few water molecules and with the guanidinium group of Arg299, a residue that is strictly conserved in all GT5 (bacterial and plant) synthases. A weaker electrostatic interaction is also observed with another conserved residue, Lys304, which in turn forms a salt bridge with an invariant glutamate at position 376. Added to these interactions, the 'open–close' transition brings the β1–α1 connecting loop of the N-terminal domain (containing the conserved motif Lys-X-Gly-Gly-Leu) into contact with the bound nucleotide. The glycine residues come into contact with the nucleotide phosphate groups and Lys15 with the ribose moiety (Figure 3B). Further experimental support to this 'closure' movement is given by the demonstration of the direct role of the Lys-X-Gly-Gly-Leu motif in nucleoside-diphospho-glucose binding that has been reported with *Escherichia coli* GS (Mahrenholz *et al*, 1988; Furukawa *et al*, 1990), although only the glycine residues but not the basic side chain appear to be essential for enzymatic activity (Furukawa *et al*, 1993). This closed conformation is indeed observed in the crystal structures of trehalose-6-P synthase (OtsA; Gibson *et al*, 2002) and other GT-B transferases.

Although the binding of ADP induces only small changes in the overall AtGS structure, it does promote important local modifications. Arg299 swings its side chain into the catalytic pocket, positioning its guanidinium group in close contact with the phosphate groups of ADP at electrostatic/hydrogen-bonding distance (2.8 Å between Arg NH2 and PO4β O3; 3.2 Å between Arg Nε and PO4α O1). Another charged residue, Lys304, also moves its side chain to approach the phosphate groups, and the aromatic ring of Tyr354 rotates 120° to stack against the adenine heterocycle. In addition, the polypeptide backbones of two loops (residues 327–331 and 374–379) and helix α14 slightly change their main-chain trace upon ADP binding (r.m.s.d.'s of 0.53, 0.74 and 0.51 Å, respectively, compared to a global r.m.s.d. of 0.31 Å for 211 equivalent Cα's in the C-terminal domain), resulting in an overall enlargement of the ADP-binding pocket.

The overall architecture of the nucleotide-binding pocket in the ADP-Glc-specific AtGS is basically the same as in OtsA, another retaining GT-B GT that preferentially binds UDP-Glc. In particular, the short connecting loop between β15 and α12 has the same length in the two proteins, and this length appears to be invariant in all GSs including both GT3 and GT5 families, as suggested by structure-weighted multiple sequence alignment. In AtGS, the central position of this loop is occupied by Tyr354, whose side chain makes stacking interactions with the adenine ring (Figure 3). Interestingly, this position is systematically occupied by an aromatic (Tyr/Phe) residue in all GSs specific for ADP-Glc (bacterial GSs and plant starch synthases), whereas promiscuous enzymes (archaea) or UDP binders (mammals, yeast) show no particular trend (a His residue is found at this position in OtsA). Also, all GT5 enzymes show two strictly conserved glycines (AtGS Gly327 and Gly329) at the end of the β14 strand, facing the nucleotide-binding pocket. This Gly-X-Gly pattern is not conserved in GT3 UDP binders and, in particular, Gly327 is replaced by a proline in OtsA.

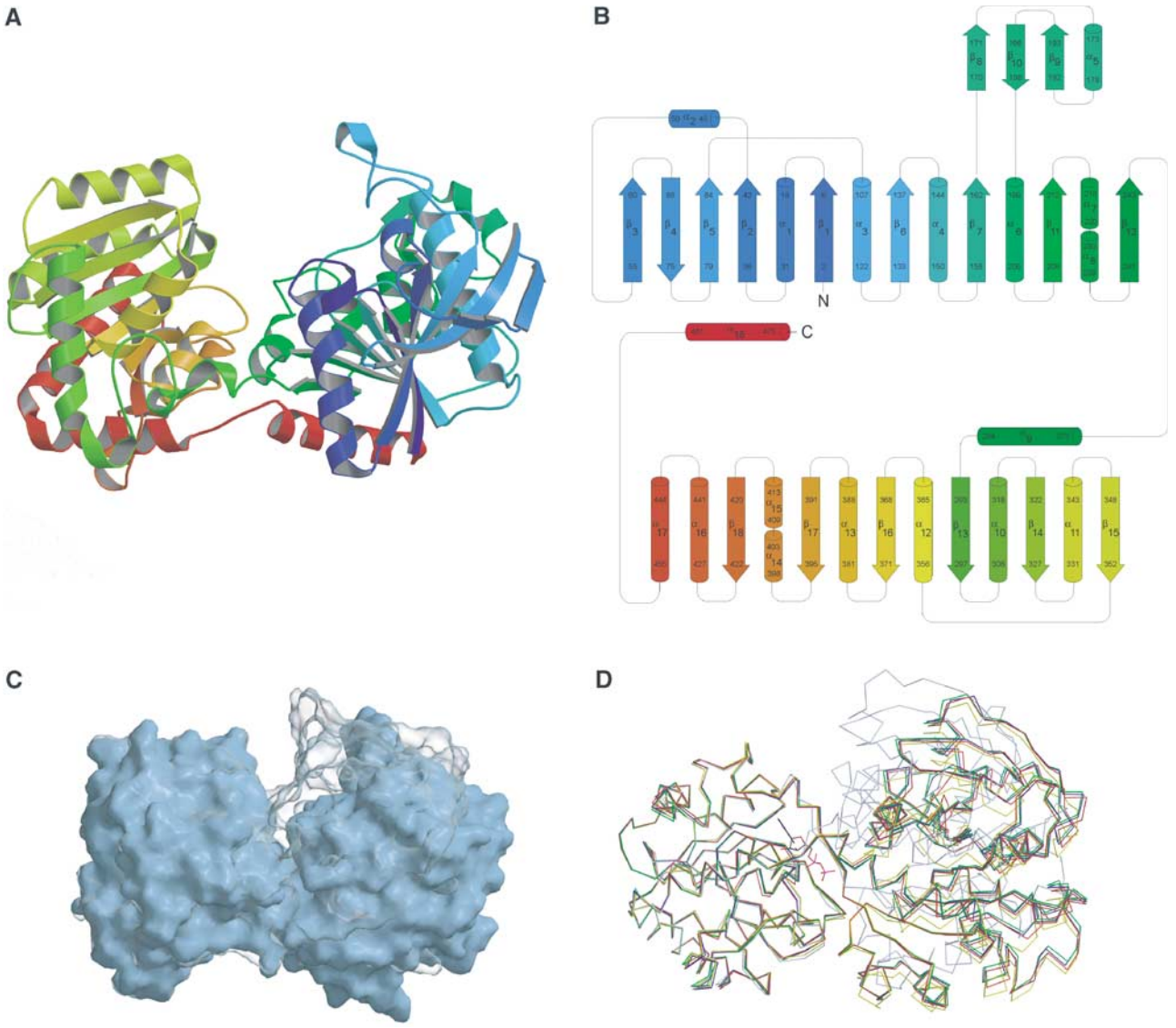


Figure 2 GS structure. (**A**) Cartoon showing the general fold and secondary structure organization of AtGS. The color ramp from blue to red indicates N- to C-terminal direction. (**B**) Cartoon representing the topology of AtGS, color-coded as in (A). Secondary structural elements, as defined by PROCHECK (Laskowski *et al*, 1993), are shown as arrows for β-strands and cylinders for α-helices comporting more than four residues. (**C**) Molecular surface representation comparing the structure of the experimental 'open' state of AtGS (solid) and the predicted 'closed' conformation (transparent). The orientation is similar to (A). (**D**) Superposition of the C-terminal domains from the four crystallographically independent AtGS molecules (unliganded: blue and green; ADP-bound: red and yellow). The predicted closed conformation is also shown in pale blue.

Structural similarity with glycogen phosphorylase

The overall structure of AtGS shows a striking topological and structural resemblance with the core of GP/MalP (GT35 family) (Watson *et al*, 1997; Rath *et al*, 2000) (Figure 4A, see also Figure 5), even though the sequence identity is very low (14% with MalP 1QM5, after structural alignment). When comparing the known GT-B enzymes, the structure of each Rossmann-fold domain of AtGS is significantly closer to those of phosphorylases from the GT35 family than to any other nucleotide-dependent GT (Figure 4B), both in terms of r.m.s.d.'s and total number of equivalent residues. This association among glycogen processing enzymes is only detectable through structural comparison, since sequence identities (10–15%) do not distinguish any particular 'superfamily' among GT-B proteins. The central β-sheet of the N-terminal domain has an identical topology in AtGS and phosphorylases, whereas one or more of its outermost strands are missing in all other known GT-B enzymes leading to some errors in the detailed prediction of GS topology when based only on sequence analyses (MacGregor, 2002). In the C-terminal domain, synthases and phosphorylases are structurally closer to each other (r.m.s.d. of 1.5–1.6 Å for ~180 equivalent residues; see Figure 4B) than they are to the other GTs (2 Å for ~100–120 residues), as a consequence of the different rearrangement of secondary structural elements. Furthermore, AtGS has two prominent insertions in the C-terminal domain, which correspond to similar features in GP but are missing in other nucleotide-dependent GT-B enzymes of known structure. The first is in the region preceding the β13 strand and includes an α-helix (α9) that superimposes

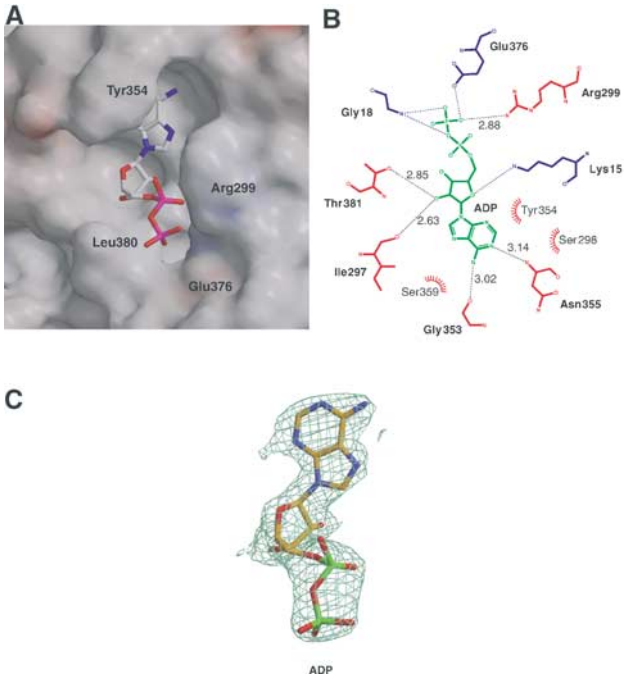


Figure 3 ADP-binding pocket. (**A**) Molecular surface representation of AtGS C-terminal domain (monomer B) showing the ADP-binding pocket. The nucleotidic ligand was reproducibly bound with higher occupancy to monomer B in the asymmetric unit; further analyses in the text hence refer only to this monomer. (**B**) Schematic representation showing the interactions between AtGS residues and ADP. Different colors distinguish between residues actually observed in the AtGS–ADP crystal complex (in red) and those residues predicted to interact after the 'closure' movement (in blue). Observed hydrogen bonding and ionic interactions are depicted with dashed lines indicating the distances in Å. Van der Waals interactions are marked with (symbols. (**C**) Electron density map (contoured at 1σ) of the bound nucleotide.

well with an equivalent helix (α17) in phosphorylases (Figure 5). The second involves the stretch of polypeptide chain including the small antiparallel sheet β8–β9–β10 (Figure 2B), which has a comparable length (25–30 residues) in both GT3 and GT5 synthases. The equivalent insertion in phosphorylases is ~20 residues longer (Figure 5) and includes all the elements defining the glycogen-storage site (Johnson *et al*, 1990), through which GP may bind to glycogen particles *in vivo*. However, several residues involved in carbohydrate contacts in the glycogen-storage site of GP are not conserved in GS and it is therefore unclear whether this exposed region, despite its similar conformation, could also fulfill a glycogen-binding role in bacterial or mammalian GS.

GSs and GPs also exhibit a strong resemblance in their catalytic and substrate-binding sites (Figure 4C). The ADP ribose and the pyridoxal groups lie in equivalent positions, as do the distal phosphate of ADP in AtGS and the inorganic phosphate in phosphorylases. Moreover, critical amino acids that interact with the glucosyl residues at positions +1 and +2 (substrate position numbering as in Davies *et al*, 1995) in MalP (PDB entries 1QM5, 1L5W and 1L6I; the numbering system for rabbit muscle GP is used for both GP and MalP throughout this manuscript) are largely conserved between the two enzyme families. These residues define an entrance tunnel in AtGS (closed form) that may bind the innermost part of the incoming glycogen chain in the same way as in MalP (Figure 4C). One of these residues is MalP Asp339, which interacts with the OH2 and OH3 groups of the sugar at position +1 and displays the same unfavorable (ϕ, ψ) angles, (60°, 180°), as the equivalent Asp138 in AtGS. An acidic residue at this position is conserved in all GT35 (Asp), GT5 (Asp) and GT3 (Glu) enzymes (Figures 4D and 5). Indeed, this Asp seems to play an important role in enzymatic activity since its mutation to Asn in the maize SSIIb starch synthase renders the enzyme inactive (Nichols *et al*, 2000). Also interacting with glucose at site +1, an invariant glutamate (Glu88) and a leucine (Leu136) are observed in phosphorylases, which correspond to conserved residues in synthases (AtGS Glu9 and Leu19; SSIIb Glu9; see Nichols *et al*, 2000). A tyrosine residue (Tyr280) that stacks against the sugar ring at position +2 is conserved in bacterial synthases (AtGS Tyr96) and is substituted by Ile/Leu in starch synthases. Facing Tyr96, the phenyl ring of Phe167 in the AtGS structure is favorably located to make stacking interactions with the second face of the sugar ring. An aromatic residue Phe/Tyr is conserved at this position in all GT5 enzymes, while an invariant leucine is found in GT3 enzymes. Interestingly, this loop corresponds to the so-called 380's loop in phosphorylases, which is in contact with the substrate chain in MalP and is known to undergo a significant conformational change upon allosteric activation in GP or upon occupation of site −1 in nonallosteric MalP.

Critical residues and their interactions in the reaction center are preserved in GP/GS (Figure 4C), strongly supporting a common catalytic mechanism. The distal phosphate of ADP in AtGS interacts with two basic residues (Arg299 and Lys304) and with Gly18, the second glycine from the conserved Lys-X-Gly-Gly-Leu motif. Interestingly, the ADP-induced movement of Arg 299 in AtGS may be equivalent to that reported for the equivalent GP residue (Arg569) when phosphorylase undergoes the allosteric T–R transition to create a high-affinity phosphate-binding site adjacent to the PLP cofactor (Barford and Johnson, 1989). The protein environment at the glucosyl donor subsite of AtGS is strikingly similar to that of phosphorylase and, to a lesser extent, to that of OtsA. AtGS residues His163 and Glu376 occupy the same location as the equivalent residues in phosphorylase (His377/Glu672; Geremia *et al*, 2002) and OtsA (His154/Asp361; Gibson *et al*, 2004), where they interact respectively with the O6 and O3 hydroxyls of the glucosyl moiety. The O6 hydroxyl also interacts with a conserved asparagine residue in phosphorylase (Asn484), which is structurally equivalent to Asn246 in AtGS. This hydrogen bond is not observed in OtsA, where O6 is engaged in a chemically similar interaction with Gln185 from a different part of the structure. All the amino-acid positions discussed above are strictly invariant in the GP/GS superfamily (with the possible exception of the Arg/Lys pair and Asn246 in GT3 enzymes), and mutagenesis evidence highlights their implication in enzymatic activity. Recent work demonstrates the direct implication of *E. coli* GS Glu377 (equivalent to AtGS Glu376) in catalysis (Yep *et al*, 2004), and the AtGS mutant His163Ala (data not shown) also proved to be completely inactive.

GT5 versus GT3 glycogen synthases

A major difference between GT3 and GT5 GSs is that only the mammalian/yeast enzymes are tightly regulated by phosphorylation and noncovalent allosteric modulation.

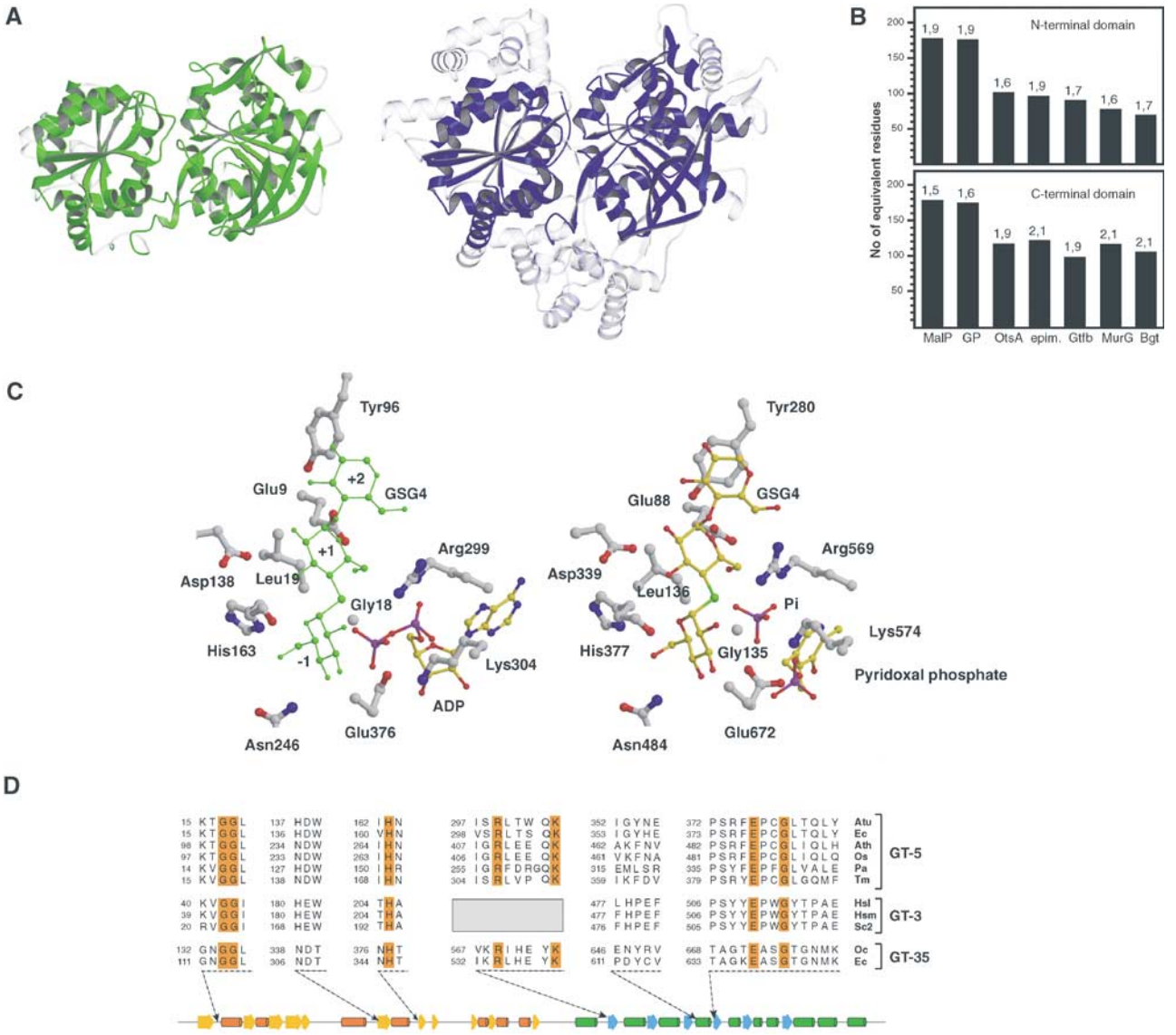
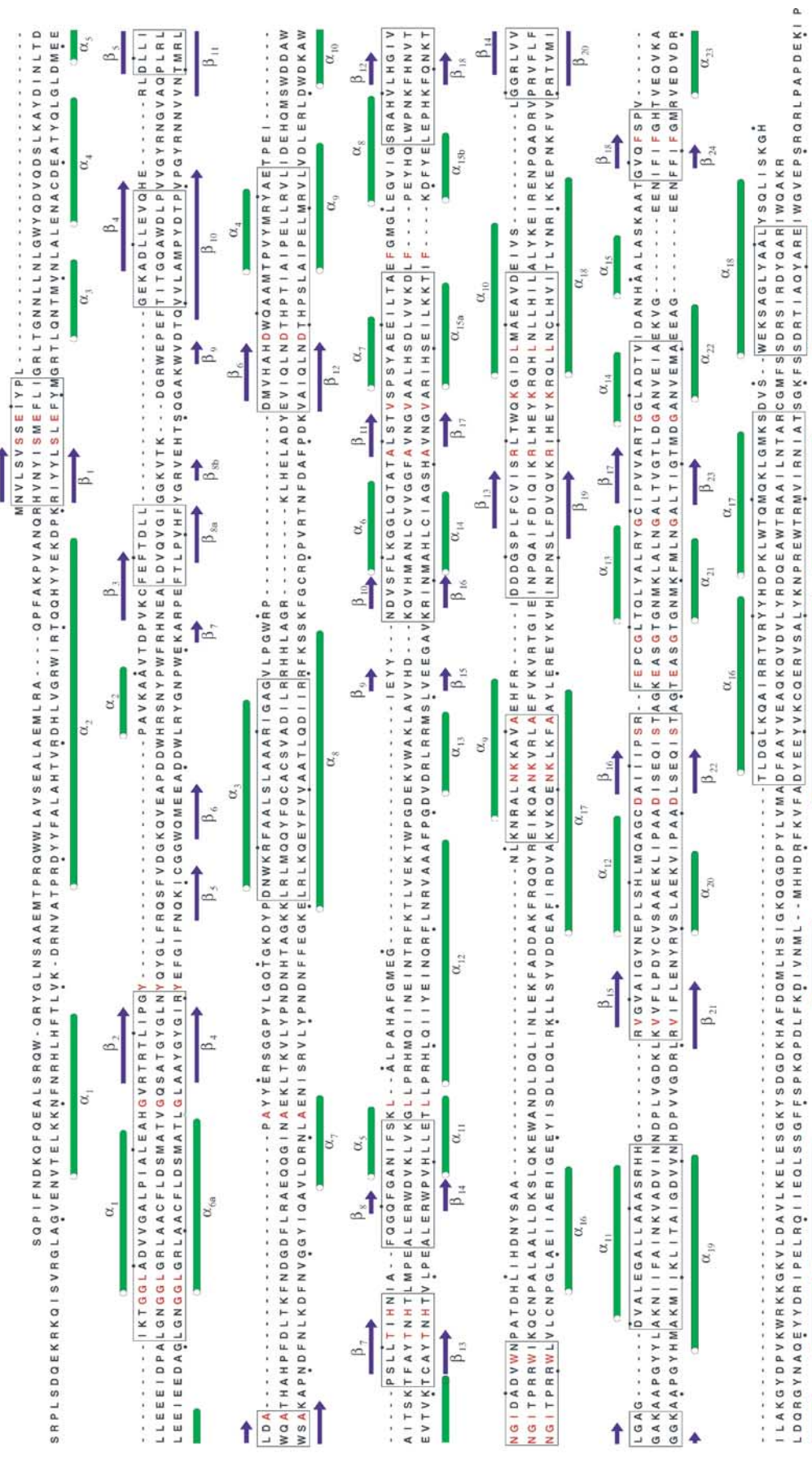


Figure 4 Homology to GP. (**A**) Cartoon representing the secondary structure organizations of AtGS (shown in green) and *E. coli* MalP (in blue) in the same orientation. After structural alignment, superimposed equivalent residues are represented in solid material, while in transparent are shown the regions that are not aligned. (**B**) Histograms showing the number of equivalent residues after pairwise structural alignment between AtGS and several GT-B enzymes: maltodextrin phosphorylase (MalP, 1QM5), glycogen phosphorylase (GP, 1A8I), trehalose-6-phosphate synthase (OtsA, 1GZ5), UDP-*N*-acetylglucosamine 2-epimerase (epim., 1F6D), UDP-glucosyl transferase (Gtfb, 1IIR), UDP-*N*-acetylglucosamine-*N*-acetylmuramyl-(pentapeptide)pyrophosphoryl-undecaprenol *N*-acetylglucosamine transferase (MurG, 1NLM) and DNA-β-glucosyltransferase (Bgt, 1JG6). The N-terminal and C-terminal domains were aligned separately. At the top of each bar, the calculated r.m.s. (in Å) of aligned Cα positions is shown. (**C**) Comparative representation of the catalytic sites of GS in its closed conformation (left view) and a ternary complex of MalP with inorganic phosphate and GSG4, a nonhydrolyzable analog of maltopentaose (right view). The sites are shown in the same orientation after structural superposition of both structures. Atoms are colored red (O), blue (N), purple (P), gray (protein C) and yellow (ligand C). The glucose acceptor oligosaccharide (as seen in MalP) is shown in green in the AtGS structure. (**D**) Multiple alignment of selected regions among members of the GS (GT3 and GT5) and GP (GT35) families. Atu: *A. tumefaciens*; Ec: *E. coli*; Ath: *Arabidopsis thaliana*; Os: *Oriza sativa*; Pa: *Pyrococcus abysii*; Tm: *Thermotoga maritima*; Hsl: *Homo sapiens* (liver); Hsm: *H. sapiens* (muscle); Sc2: *Saccharomyces cerevisiae* (Gys2); Oc: *Oryctolagus cuniculus*. Nonaligned regions for GT3 enzymes are depicted with a gray box. The secondary structure of AtGS is shown as a reference at the bottom.

According to multiple sequence alignments among GT3, GT5 and GT35 proteins, weighted by the structural alignment of AtGS versus GT35, the regulatory phosphorylation sites in human GS are probably located in the 25-residue N-terminal and 120-residue C-terminal insertions, outside the nonregulated GT-B core of bacterial GSs (Figure 6). This is consistent with the susceptibility to proteolysis and concomitant loss of regulation of rabbit liver and yeast GSs (Camici *et al*, 1982; Hardy and Roach, 1993), together with the demonstration of *in vivo* phosphorylation sites in yeast (Hardy and Roach, 1993) and mammalian (Skurat and Roach, 1995) GSs at specific Ser residues (one Thr in yeast) located in the C-terminal extension, as well as two N-terminal serines for mammalian GS. Mutagenesis studies of the yeast (Pederson *et al*, 2000) and mammalian (Hanashiro and Roach, 2002) enzymes identified two conserved Arg clusters, one in the loop between α16 and α17 helices (Arg579/580/582) and the other on α17 itself (Arg586/588/591), that are part of



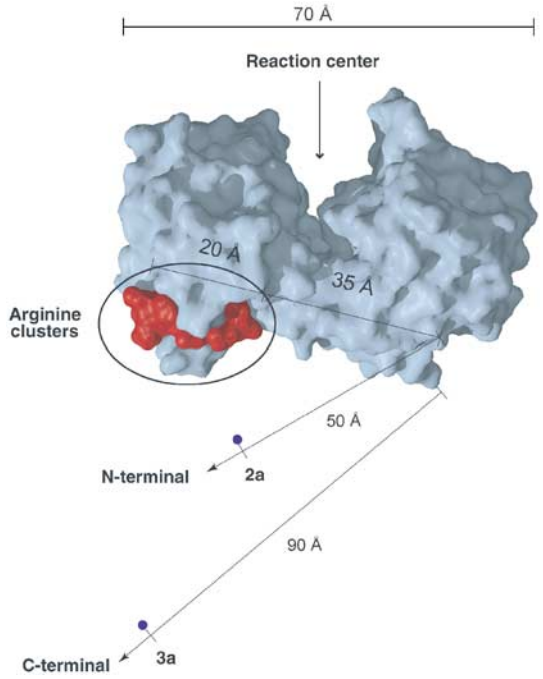


Figure 6 Molecular surface representation of the GS core, showing the equivalent position of the arginine clusters in the mammalian/yeast (GT3) allosteric site (in red) with respect to the active center. Assuming an extended main-chain conformation, approximate distances are shown for two relevant phosphorylation sites, one in the N-terminal (2a) and the other in the C-terminal (3a) extensions of GT3 enzymes.

the intra- and/or intermolecular allosteric site(s) for both the phospho-Ser/Thr residues and the noncovalent activator glucose-6-phosphate (Glc-6-P). The equivalent residues in AtGS are located at one end of the C-terminal domain, far from the reaction center (Figure 6). Binding of the phosphorylated residues and/or the Glc-6-P activator can therefore propagate the allosteric effects to the active site either by tertiary or quaternary structure rearrangements (for instance, through the control of the open–close transition, modulating the enzyme 'switch' to its active state) or by intramolecular changes connecting the allosteric site(s) to the reaction center. Mutagenesis evidence in favor of a long-distance effect is surprisingly provided by the Lys277Gln replacement in *E. coli* GS (Furukawa *et al*, 1994), which resulted in a 140-fold reduction of k_{cat} and led the authors to propose a direct involvement of this residue in catalysis. In fact, the AtGS structure shows that the lysine side chain (strictly conserved in GT5 and GT35 enzymes) is located far away (~30 Å) from the active center. Its side chain is hydrogen bonded to Tyr441 at the C-terminal end of α16 (corresponding to the position of the GT3 Arg cluster) and to Asp367 at the α12–β16 connecting loop, which in GT3 enzymes includes Arg498 (also shown to be involved in Glc-6-P binding; Pederson *et al*, 2000). Although indirect, these results do support a long-distance effect, and the equivalent location of GT5 Lys277 and the GT3 Arg cluster is, at the least, intriguing.

Materials and methods

Protein expression and crystallization

Protein production, purification and crystallization of *A. tumefaciens* GS were performed as described (Guerin *et al*, 2003; Ugalde *et al*, 2003).

Data collection and phasing

Single crystals of the SeMet-labeled protein (10 Se atoms per molecule) were drawn out of the crystallization drops in cryoloops (Hampton Research) and frozen in liquid nitrogen in the presence of a cryoprotectant solution (the same mother liquor except for 15% PEG 4000 and 30% glycerol). Complete X-ray diffraction data sets were collected at 110 K using synchrotron radiation (beamline ID29, European Synchrotron Radiation Facility, Grenoble) with an a ADSC Q210 detector. Bragg reflection intensities were integrated with MOSFLM6.2.0, merged with SCALA2.7.5 and reduced to structure factor amplitudes with TRUNCATE4.1 (Collaborative Computational Project, 1994). The structure was determined by a single-wavelength anomalous diffraction experiment. Of the 20 Se sites in the asymmetric unit, 19 were found by direct methods with the program Shake'n'Bake (Weeks and Miller, 1999) using the anomalous difference intensities. The positions, *B*-factors and occupancies of the anomalous scatterers were refined with the program Sharp (Bricogne *et al*, 2003), with all 20 Se atoms identified after three rounds of refinement and inspection of log-likelihood gradient residual Fourier maps. Density modification was performed with DM (Collaborative Computational Project, 1994), resulting in a readily interpretable map.

Model building, refinement and validation

Model building was performed with O8.0 (Jones *et al*, 1991) alternated with reciprocal space refinement cycles. Restrained refinement was carried out with the program CNS1.1 (Brünger *et al*, 1998) for the unliganded AtGS structure, using the anomalous data (Friedel pairs considered as independent reflections) and the measured values of $\Delta f'$ and $\Delta f''$ to correct the scattering factors of the Se atoms. A fraction (5%) of the data set was set aside for R_{free} calculation. Tight NCS restraints were initially imposed, but were gradually relaxed and completely removed during the final refinement stages.

Because of significant unit cell changes, molecular replacement methods were used to determine the structure of the AtGS–ADP complex. In this case, initial refinement was performed with CNS, and toward the end of refinement Refmac5 (Collaborative Computational Project, 1994) was used to perform TLS refinement assuming two rigid bodies (the N- and C-terminal domains) per molecule. The final models display a good stereochemistry as evaluated with the program PROCHECK (Collaborative Computational Project, 1994), with less than 0.3% of the residues in disallowed regions of the Ramachandran plots.

Modeling of the 'closed' state of AtGS

Two domains were defined in AtGS in order to maximize the superposition with MalP (PDB code 1QM5). The N-terminal domain of AtGS includes residues 1–244 and 457–477, while the C-terminal one extends through residues 245–456. An overall interdomain rotation of ~24° was found to optimize the separate superpositions of both AtGS domains to MalP. The global r.m.s.d.'s were 2.28 Å for 254 equivalent Cα atoms before optimization and 1.87 Å for 360 residues after applying the rigid-body rotation. Atomic coordinates were then generated for hydrogen atoms and the model was subjected to energy minimization using the CHARMM force field as implemented in CNS (Brünger *et al*, 1998).

Sequence alignments

Multiple sequence alignments of GSs (families GT3 and GT5) and GPs (family GT35) were carried out with the program CLUSTALW v1.83 (Thompson *et al*, 1994). Combined structural and multiple sequence alignments were performed using the program Indonesia

Figure 5 Structural alignment of AtGS, *E. coli* MalP and rabbit muscle GP. Structurally equivalent regions are boxed, conserved residues are shown in red. Secondary structural elements (α-helices and β-strands) are shown above the sequence for AtGS and below the sequence for GP. Every tenth residue in the AtGS and GP sequences is dotted.

(Madsen, Johansson and Kleywegt, unpublished results; see http://alpha2.bmc.uu.se/dennis for details).

Acknowledgements

We thank Andrea Merás IIB-UNSAM-CONICET for technical assistance during the purification of GS. RAU is a member of the research career of CONICET (Argentina). JEU is a fellow of CONICET (Argentina). We thank the Institut Pasteur (France), CONICET (Argentina) and Agencia Nacional de Promocion Cientifica y Tecnologica ANPCyT (Argentina), PICT 01-6565, for financial support.

Atomic coordinates and structure factors have been deposited with the Protein Data Bank, accession codes 1RZU and 1RZV.

References

Abad MC, Binderup K, Rios Steiner J, Arni RK, Preiss J, Geiger JH (2002) The X-ray crystallographic structure of *Escherichia coli* branching enzyme. *J Biol Chem* **277:** 42164–42170

Aschenbach WG, Suzuki Y, Breeden K, Prats C, Hirshman MF, Dufresne SD, Sakamoto K, Vilardo PG, Steele M, Kim JH, Jing SL, Goodyear LJ, DePaoli-Roach AA (2001) The muscle-specific protein phosphatase PP1G/R(GL)(G(M)) is essential for activation of glycogen synthase by exercise. *J Biol Chem* **276:** 39959–39967

Azpiazu I, Manchester J, Skurat AV, Roach PJ, Lawrence JC (2000) Control of glycogen synthesis is shared between glucose transport and glycogen synthase in skeletal muscle fibers. *Am J Physiol Endocrinol Metab* **278:** E234–E243

Ball S, Morell MK (2003) From bacterial glycogen to starch: understanding the biogenesis of the plant starch granule. *Annu Rev Plant Biol* **54:** 207–233

Barford D, Johnson LN (1989) The allosteric transition of glycogen phosphorylase. *Nature* **340:** 609–616

Bricogne G, Vonrhein C, Flensburg C, Schiltz M, Paciorek W (2003) Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta Crystallogr D* **59:** 2023–2030

Brünger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang J-S, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr D* **54:** 905–921

Buchbinder JL, Rath VL, Fletterick RJ (2001) Structural relationships among regulated and unregulated phosphorylases. *Annu Rev Biophys Biomol Struct* **30:** 191–209

Camici M, DePaoli-Roach AA, Roach PJ (1982) Rabbit liver glycogen synthase. Susceptibility of the enzyme subunit to proteolysis. *J Biol Chem* **257:** 9898–9901

Charnock SJ, Davies GJ (1999) Structure of the nucleotide-diphospho-sugar transferase, SpsA from *Bacillus subtilis*, in native and nucleotide-complexed forms. *Biochemistry* **38:** 6380–6385

Collaborative Computational Project Number 4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D* **50:** 760–763

Coutinho PM, Deleury E, Davies GJ, Henrissat B (2003) An evolving hierarchical family classification for glycosyltransferases. *J Mol Biol* **328:** 307–317

Davies GJ, Tolley SP, Henrissat B, Hjort C, Schulein M (1995) Structures of oligosaccharide-bound forms of the endoglucanase V from *Humicola insolens* at 1.9 Å resolution. *Biochemistry* **34:** 16210–16220

Furukawa K, Tagaya M, Inouye M, Preiss J, Fukui T (1990) Identification of lysine 15 at the active site in *Escherichia coli* glycogen synthase. Conservation of Lys-X-Gly-Gly sequence in the bacterial and mammalian enzymes. *J Biol Chem* **265:** 2086–2090

Furukawa K, Tagaya M, Tanizawa K, Fukui T (1993) Role of the conserved Lys-X-Gly-Gly sequence at the ADP-glucose-binding site in *Escherichia coli* glycogen synthase. *J Biol Chem* **268:** 23837–23842

Furukawa K, Tagaya M, Tanizawa K, Fukui T (1994) Identification of Lys277 at the active site of *Escherichia coli* glycogen synthase. Application of affinity labeling combined with site-directed mutagenesis. *J Biol Chem* **269:** 868–871

Geremia S, Campagnolo M, Schinzel R, Johnson LN (2002) Enzymatic catalysis in crystals of *Escherichia coli* maltodextrin phosphorylase. *J Mol Biol* **322:** 413–423

Gibbons BJ, Roach PJ, Hurley TD (2002) Crystal structure of the autocatalytic initiator of glycogen biosynthesis, glycogenin. *J Mol Biol* **319:** 463–477

Gibson RP, Tarling CA, Roberts S, Withers SG, Davies GJ (2004) The donor subsite of trehalose-6-phosphate synthase: binary complexes with UDP-glucose and UDP-2-deoxy-2-fluoro-glucose at 2 Å resolution. *J Biol Chem* **279:** 1950–1955

Gibson RP, Turkenburg JP, Charnock SJ, Lloyd R, Davies GJ (2002) Insights into trehalose synthesis provided by the structure of the retaining glucosyltransferase OtsA. *Chem Biol* **9:** 1337–1346

Guerin ME, Buschiazzo A, Ugalde JE, Ugalde RA, Alzari PM (2003) Preliminary crystallographic studies of glycogen synthase from *Agrobacterium tumefaciens*. *Acta Crystallogr D* **59:** 526–528

Hanashiro I, Roach PJ (2002) Mutations of muscle glycogen synthase that disable activation by glucose 6-phosphate. *Arch Biochem Biophys* **397:** 286–292

Hardy TA, Roach PJ (1993) Control of yeast glycogen synthase-2 by COOH-terminal phosphorylation. *J Biol Chem* **268:** 23799–23805

Henrissat B, Deleury E, Coutinho PM (2002) Glycogen metabolism loss: a common marker of parasitic behaviour in bacteria? *Trends Genet* **18:** 437–440

Holm L, Sander C (1995) Evolutionary link between glycogen phosphorylase and a DNA modifying enzyme. *EMBO J* **14:** 1287–1293

Hondoh H, Kuriki T, Matsuura Y (2003) Three-dimensional structure and substrate binding of *Bacillus stearothermophilus* neopullulanase. *J Mol Biol* **326:** 177–188

Hu Y, Chen L, Ha S, Gross B, Falcone B, Walker D, Mokhtarzadeh M, Walker S (2003) Crystal structure of the MurG:UDP-GlcNAc complex reveals common structural principles of a superfamily of glycosyltransferases. *Proc Natl Acad Sci USA* **100:** 845–849

Johnson LN (1992) Glycogen phosphorylase: control by phosphorylation and allosteric effectors. *FASEB J* **6:** 2274–2282

Johnson LN, Acharya KR, Jordan MD, McLaughlin PJ (1990) Refined crystal structure of the phosphorylase–heptulose 2-phosphate–oligosaccharide–AMP complex. *J Mol Biol* **211:** 645–661

Jones TA, Zou JY, Cowan SW, Kjeldgaard M (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr D* **47:** 110–119

Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr* **26:** 283–291

Leloir LF, Cardini CE (1957) Biosynthesis of glycogen from uridine diphosphate glucose. *J Am Chem Soc* **79:** 6340–6341

MacGregor EA (2002) Possible structure and active site residues of starch, glycogen, and sucrose synthases. *J Prot Chem* **21:** 297–306

Mahrenholz AM, Wang YH, Roach PJ (1988) Catalytic site of rabbit glycogen synthase isozymes. Identification of an active site lysine close to the amino terminus of the subunit. *J Biol Chem* **263:** 10561–10567

Manchester J, Skurat AV, Roach P, Hauschka SD, Lawrence JC (1996) Increased glycogen accumulation in transgenic mice overexpressing glycogen synthase in skeletal muscle. *Proc Natl Acad Sci USA* **93:** 10707–10711

Morera S, Lariviere L, Kurzeck J, Aschke-Sonnenborn U, Freemont PS, Janin J, Ruger W (2001) High resolution crystal structures of T4 phage beta-glucosyltransferase: induced fit and effect of substrate and metal binding. *J Mol Biol* **311:** 569–577

Nichols DJ, Keeling PL, Spalding M, Guan H (2000) Involvement of conserved aspartate and glutamate residues in the catalysis and substrate binding of maize starch synthase. *Biochemistry* **39:** 7820–7825

Pederson BA, Cheng C, Wilson WA, Roach PJ (2000) Regulation of glycogen synthase. Identification of residues involved in regulation by the allosteric ligand glucose-6-P and by phosphorylation. *J Biol Chem* **275:** 27753–27761

Rath VL, Ammirati M, LeMotte PK, Fennell KF, Mansour MN, Danley DE, Hynes TR, Schulte GK, Wasilko DJ, Pandit J (2000) Activation of human liver glycogen phosphorylase by alteration of the secondary structure and packing of the catalytic core. *Mol Cell* **6:** 139–148

Roach PJ (2002) Glycogen and its metabolism. *Curr Mol Med* **2:** 101–120

Saltiel AR (2001) New perspectives into the molecular pathogenesis and treatment of type 2 diabetes. *Cell* **104:** 517–529

Skurat AV, Roach PJ (1995) Phosphorylation of sites 3a and 3b (Ser640 and Ser644) in the control of rabbit muscle glycogen synthase. *J Biol Chem* **270:** 12491–12497

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22:** 4673–4680

Ugalde JE, Parodi AJ, Ugalde RA (2003) *De novo* synthesis of bacterial glycogen: *Agrobacterium tumefaciens* glycogen synthase is involved in glucan initiation and elongation. *Proc Natl Acad Sci USA* **100:** 10659–10663

Vrielink A, Ruger W, Driessen HP, Freemont PS (1994) Crystal structure of the DNA modifying enzyme beta-glucosyltransferase in the presence and absence of the substrate uridine diphosphoglucose. *EMBO J* **13:** 3413–3422

Watson KA, Schinzel R, Palm D, Johnson LN (1997) The crystal structure of *Escherichia coli* maltodextrin phosphorylase provides an explanation for the activity without control in this basic archetype of a phosphorylase. *EMBO J* **16:** 1–14

Weeks CM, Miller R (1999) Optimizing Shake-and-Bake for proteins. *Acta Crystallogr D* **55:** 492–500

Wrabl JO, Grishin NV (2001) Homology between O-linked GlcNAc transferases and proteins of the glycogen phosphorylase superfamily. *J Mol Biol* **314:** 365–374

Yep A, Ballicora MA, Sivak MN, Preiss J (2004) Identification and characterization of a critical region in the glycogen synthase from *Escherichia coli*. *J Biol Chem* **279:** 8359–8367